

When should clinical guidelines be updated?

Paul Shekelle, Martin P Eccles, Jeremy M Grimshaw, Steven H Woolf

Considerable resources are being expended internationally on the development of clinical practice guidelines.¹ Although consensus is increasing about methods for developing evidence based guidelines,² less attention has been paid to the process for assessing when guidelines should be updated. The most common advice is for guidelines to include a scheduled review date. This could result in wasted resources, however, if a full update is undertaken prematurely in a slowly evolving field, or in guidelines in a rapidly evolving field becoming out of date before the scheduled review. Some guidelines state that they should be updated when new information becomes available. It is unclear, however, how this should be done, and we are unaware of any systematic attempts to devise a method. In this paper we propose a set of principles and a pragmatic model for assessing whether guidelines need to be updated.

What situations might require clinical guidelines to be updated?

Changes in evidence on the existing benefits and harms of interventions

New information about the magnitude of benefits and harms may make the pre-existing guideline invalid. The surgical risk of carotid endarterectomy, for example, has fallen substantially over the past 30 years, altering the risk-benefit ratio in favour of performing the operation for selected patients with symptomatic, high grade carotid stenosis.³⁻⁵

Changes in outcomes considered important

New evidence may identify as important outcomes that were previously unappreciated or wholly unrecognized. Quality of life, for example, an end point often not considered in earlier research and guidelines, is receiving increasing recognition as an important outcome of health care.

Changes in available interventions

Since the development of a guideline, new preventive, diagnostic, or treatment interventions may have emerged to complement or supersede other interventions. A guideline on unstable angina, for example, would need to reflect the new role of coronary artery stents and glycoprotein IIb/IIIa in improving outcomes.^{6,7}

Changes in evidence that current practice is optimal

Guidelines are developed to help narrow the gap between ideal and current clinical practice. This gap could narrow over time to the point that a guideline is no longer needed. For example, a national survey of surgical specialties in Scotland two years after the dissemination of a national guideline found that 90% of patients appropriately received deep vein thromboprophylaxis.⁸

Changes in values placed on outcomes

The values that individuals or society place on different outcomes may change over time. Economic issues, for

Summary points

Changes in evidence, the values placed on evidence, the resources available for health care, and improvements in current performance are all possible reasons for updating clinical guidelines

The need for an efficient mechanism for identifying when guidelines require updating is urgent

A possible model for assessing validity of guidelines is based on a combination of multidisciplinary expert opinion and limited literature searches

example, have received little attention in most guidelines but will be considered explicitly in guidelines developed by the UK National Institute for Clinical Excellence.

Changes in resources available for health care

Guidelines may need to be updated to permit increased delivery of services if the level of available resources increases over time. The recent expiry of the patent on fluoxetine, for example, which is expected to reduce its price through competition, may influence guidelines for antidepressant drugs.⁹

Model for assessing whether a guideline needs updating

How can we assess whether there have been sufficient changes in these factors to warrant updating a guideline? We focus here on changes in evidence or performance. Changes in the values placed on outcomes often reflect societal norms. Measuring the values placed on outcomes and how these change over time is complex and not dealt with here. When changes occur in the availability of resources for health care or the costs of interventions, a generic policy on updating is unlikely to be helpful, because policy-makers in disparate healthcare systems consider different factors in deciding whether services remain affordable.

We therefore focus on defining when new information on interventions, outcomes, and performance justifies updating guidelines. This process includes two stages: identifying important new evidence and assessing whether the new evidence warrants updating. Ideally, the best way to identify important new evidence would be to conduct a systematic review, but this would be costly and time consuming. It would be tantamount to completing the first step of updating, rather than determining whether updating was necessary; a more timely and efficient screening process is needed.

Greater Los Angeles Veterans Affairs Healthcare System, Los Angeles, CA 90073, USA

Paul Shekelle
senior research associate

Centre for Health Services Research, University of Newcastle upon Tyne NE2 4AA

Martin P Eccles
professor of clinical effectiveness

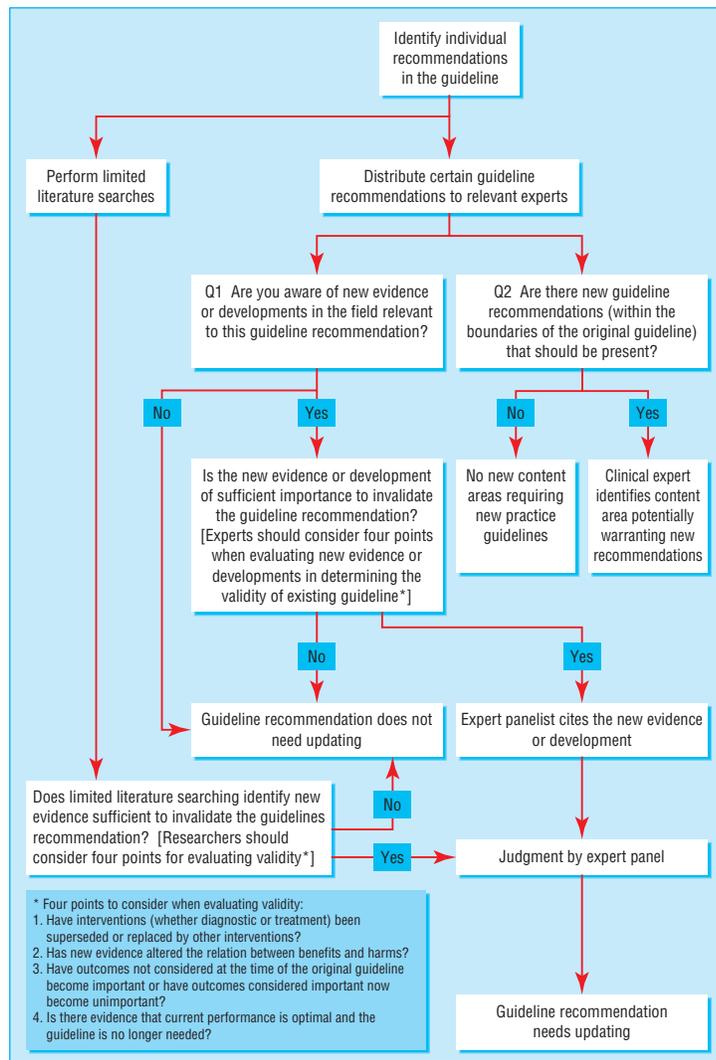
Health Services Research Unit, University of Aberdeen, Aberdeen AB25 2ZD

Jeremy M Grimshaw
professor of health services research

Department of Family Practice, Virginia Commonwealth University, Fairfax, VA 22033, USA
Steven H Woolf
professor of family medicine

Correspondence to: P Shekelle, RAND, 1700 Main Street, PO Box 2138, M-26 Santa Monica, CA 90407-2138, USA
shekelle@rand.org

BMJ 2001;323:155-7



Proposed model for assessing the current validity of guidelines

We posit that evidence sufficient to invalidate an existing guideline would, in general, be known to experts in the field or have been published as important articles in major general interest or specialty journals. We therefore advocate a model based on expert opinion and focused literature reviews to assess when guidelines need updating (figure).

Our model proposes that a multidisciplinary group of experts reviews selected recommendations within the guideline. Potential experts for this task could be recruited from the original guideline development group, complemented by additional topic experts and generalists with expertise in critical appraisal. These experts would be asked whether they were aware of new evidence or developments in the field relevant to the guideline recommendation and, if so, whether this evidence was sufficient to invalidate the guideline recommendation. This judgment of “sufficiency” should be based on the criteria presented above (new interventions, new data on benefits and harms, new outcomes, or evidence that the guideline is no longer needed).

We propose that the experts should also be asked to identify any changes in the interventions available—for example, new or outmoded measures—that might

invalidate the recommendation. Experts should be asked to provide references to support their views regarding new evidence or interventions.

This process would be supplemented by literature searches to reduce the chance of oversights by experts. The search would focus on major general interest and specialty medical journals, timed from when the literature search for the original guideline ended. The searches could initially target review articles, editorials, and commentaries (“sentinel” markers of new evidence sufficient to change practice), new guidelines on the topic in current registries (www.guidelines.gov for example), and articles that reference the previous practice guideline or major studies (such as the Science Citation index). These and other search methods should undergo formal comparison to weigh the relative accuracy and expediency of searches versus consultation with experts in finding important new evidence.

Judging when it is appropriate to retain a guideline

The next step is an independent assessment of whether the new evidence or interventions identified are of sufficient importance to invalidate the guideline recommendation. In some cases the new information will provide prima facie evidence that the guideline recommendation is invalid—for example, if a large clinical trial shows convincingly that a recommended treatment is ineffective or harmful. For other situations, however, this assessment will necessarily require judgment, and we think such judgments are generally more balanced if they involve both topic experts and generalists with expertise in guideline development.

Within any individual guideline there will be some recommendations that are invalid while others remain current. A guideline on congestive heart failure, for example, includes 37 individual recommendations.¹⁰ How many must be invalid to require updating the whole guideline? Clearly a guideline needs updating if most recommendations are out of date, with new evidence showing that the recommended interventions are inappropriate, ineffective, or superseded by new interventions. But in other cases a single, outdated recommendation could invalidate the entire document. Judgments about the whether a guideline needs updating are inherently subjective and reflect the clinical importance and number of invalid recommendations.

Future development of guidelines

Future efforts in the development of guidelines could consider performing the type of work we propose here as a continuous prospective exercise. Adding a mechanism for clinicians to provide feedback on when updating is necessary—such as a website and electronic formats that permit excision of outdated recommendations—could also prove useful.

PS is a senior research associate of the Veterans Affairs Health Services Research and Development Service. The opinions in the article are those of the authors and do not necessarily represent the opinions of the Agency for Healthcare Research and Quality, the US Department of Health and Human Services, the US Department of Veterans Affairs, the Scottish Executive Department of Health, or the MRC Health Services Research Collaboration.

Contributors: PS had the original idea and will act as guarantor; all four authors contributed to the discussions and to the writing of the article.

Funding: PS is supported by a contract from the Agency for Healthcare Research and Quality to the Southern California Evidence-based Practice Center. The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Executive Department of Health. The Health Services Research Unit and Centre for Health Services Research are members of the MRC Health Services Research Collaboration.

Competing interests: JMG is a member of the Guidelines Advisory Committee for the National Institute for Clinical Excellence and the methodological adviser to the Scottish Intercollegiate Guidelines Network. MPE is chairman of the Guidelines Advisory Committee for the National Institute for Clinical Excellence. SHW is a member of the US Preventive Services Task Force and other practice guideline panels involved in updating.

- 1 Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: Potential benefits, limitations, and harms of clinical guidelines. *BMJ* 1999;318:527-30.
- 2 Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines. Developing guidelines. *BMJ* 1999;318:593-6.
- 3 Fields WS, Maslenikov V, Meyer JS, Hass WK, Remington RD, Macdonald M. Joint study of extracranial arterial occlusion. V: Progress report of prognosis following surgery or nonsurgical treatment for transient cerebral ischemic attacks and cervical carotid artery lesions. *JAMA* 1970;211:1993-2003.
- 4 North American Symptomatic Carotid Endarterectomy Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N Engl J Med* 1991;325:445-53.

- 5 European Carotid Surgery Trialists' Collaborative Group. MRC European Carotid Surgery Trial. Interim results for symptomatic patients with severe (70-99%) or with mild (0-29%) carotid stenosis. *Lancet* 1991;337:1235-43.
- 6 Platelet Receptor Inhibition in Ischemic Syndrome Management in Patients Limited by Unstable Signs and Symptoms (PRISM-PLUS) Study Investigators. Inhibition of the platelet glycoprotein IIb/IIIa receptor with tirofiban in unstable angina and non-Q-wave myocardial infarction. *N Engl J Med* 1998;338:1488-97.
- 7 Lincoff AM, Califf RM, Moliterno DJ, Ellis SG, Ducas J, Kramer JH, et al. Complementary clinical benefits of coronary-artery stenting and blockade of platelet glycoprotein IIb/IIIa receptors. Evaluation of Platelet IIb/IIIa Inhibition in Stenting Investigators. *N Engl J Med* 1999;341:3109-27.
- 8 Campbell SE, Walker AE, Grimshaw JM, Campbell MK, Lowe GDO, the TEMPEST Group, et al. The prevalence of prophylaxis for venous thromboembolism in acute hospital trusts [abstract]. *J Epidemiol Community Health* 1999;53:669.
- 9 Eccles M, Freemantle N, Mason J. North of England evidence-based guideline development project: summary version of guidelines for the choice of antidepressants for depression in primary care. North of England Anti-depressant Guideline Development Group. *Fam Pract* 1999;16:103-11.
- 10 Konstam MA, Dracup K, Baker DW, Bottoff MB, Brock NH, Dacey RA, et al. *Heart failure: evaluation and care of patients with left-ventricular systolic dysfunction. Clinical practice guideline No 11.* Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services, 1994. (AHCPR publication No 94-0612.)

(Accepted 12 March 2001)

Systematic reviews in health care

Systematic reviews of evaluations of diagnostic and screening tests

Jonathan J Deeks

Tests are routinely used in medicine to screen for, diagnose, grade, and monitor the progression of disease. Diagnostic information is obtained from a multitude of sources, including imaging and biochemical technologies, pathological and psychological investigations, and signs and symptoms elicited during history taking and clinical examinations.¹ Each of these items of information can be regarded as a result of a separate diagnostic or screening "test." Systematic reviews of evaluations of tests are undertaken for the same reasons as systematic reviews of treatment interventions: to produce estimates of test performance and impact based on all available evidence, to evaluate the quality of published studies, and to account for variation in findings between studies.²⁻⁵ Reviews of studies of diagnostic accuracy involve the same key stages of defining questions, searching the literature, evaluating studies for eligibility and quality, and extracting and synthesising data. However, studies that evaluate the accuracy of tests have a unique design requiring different criteria to appropriately assess the quality of studies and the potential for bias. Additionally, each study reports a pair of related summary statistics (for example, sensitivity and specificity) rather than a single statistic (such as a risk ratio) and hence requires different statistical methods to pool the results of the studies. This article concentrates on the dimensions of study quality and the advantages and disadvantages of different summary statistics for combining studies in meta-analysis. Other aspects,

Summary points

Systematic reviews of studies of diagnostic accuracy differ from other systematic reviews in the assessment of study quality and the statistical methods used to combine results

Important aspects of study quality include the selection of a clinically relevant cohort, the consistent use of a single good reference standard, and the blinding of results of experimental and reference tests

The choice of statistical method for pooling results depends on the summary statistic and sources of heterogeneity, notably variation in diagnostic thresholds

Sensitivities, specificities, and likelihood ratios may be combined directly if study results are reasonably homogeneous

When a threshold effect exists, study results may be best summarised as a summary receiver operating characteristic curve, which is difficult to interpret and apply to practice

including searching the literature and further technical details, are discussed elsewhere.⁶

This is the third in a series of four articles

Imperial Cancer Research Fund/
NHS Centre for Statistics in
Medicine, Institute
of Health Sciences,
Oxford OX3 7LF
Jonathan J Deeks
senior medical
statistician

Correspondence to:
J J Deeks
J.Deeks@icrf.
icnet.uk

Series editor:
Matthias Egger

BMJ 2001;323:157-62